Remarks

The Examiner rejected claims 4, 7-10, 11-14, 15 and 16 under 35 U.S.C. 101 as being allegedly directed to non-statutory subject matter. In response, as suggested by the Examiner, applicant has amended base claims 4, 7 and 11 to recite a "processor" therein. In addition, base claim 15 has been amended to incorporate the Examiner's suggested language in its preamble. As such, claims 4, 7, 11 and 15 as amended, together with their dependent claims, are patentable under 35 U.S.C. 101.

The Examiner also rejected claims 1-21 under 35 U.S.C. 103(a) as being allegedly obvious over Kane in view of Seilhamer. In response, base claims 1, 4, 7, 11, 15 and 18 have been amended. Claims 3, 13, 17 have also been amended to properly reference amended base claims 1, 11 and 15, respectively.

In many instances, an organization receiving data records from multiple data providers finds that each data provider formats its data records differently. This can create a significant challenge for the organization which must both reformat each record received from the various sources and identify and eliminate repetitive data records. For example, an entity such as an information assistance service provider, which receives from various telecommunications companies (e.g., AT&T, SBC, etc.) data records containing customers' addresses, telephone numbers, etc., typically must incorporate all of the data received into its database(s) while ensuring that repetitive data is eliminated. When the data records received from the telecommunication companies are in different formats (as is typical), the task of combining the data only becomes greater.

The invention is directed to a system and method for incorporating data received from different sources into one or more databases having records in a specified format. Data (e.g., records containing names and telephone numbers) are received from multiple sources (p. 6, lines 1-3). In accordance with the invention, a converter routine converts each received record into a data record conforming to a uniform format, referred to as a "uniform data record" (p. 6, lines 7-10). For example, the uniform format may require, say, that each record comprise a first field containing an identifier, a second field containing a telephone number, a third field containing an

address, etc. (Figs. 4C-4D). A normalizer then modifies the contents of the fields within each record, as necessary, to conform to a predetermined nomenclature. Thus, data in the address field of one record containing the letters "California" may be changed to "CA," while data in another record containing the letters "Ca" may also be changed to "CA" (p. 9, lines 15-20).

It should be noted that the configuration of normalizer routine 110 may change from time to time. Thus, in accordance with an aspect of the invention, the time (including, e.g., date) when a uniform data record was normalized is registered in association with the record (e.g., Last Normalized Date 206). The registered time serves as an indicator of what configuration version of the normalizer routine was applied to the record. When there is a need to change the normalization process and to renormalize the uniform data, the uniform data records having a last normalized date that predates the current normalized date can be readily identified based on their associated registered times. The records as identified may then be renormalized using the current version of the normalization routine. P. 13, lines 6-19.

After the received data records are reformatted and normalized, a processor identifies sets of potentially equivalent normalized data records (p.14, lines 11-15). For example, two normalized data records containing identical telephone numbers (in their respective telephone number fields) may be identified as records that may represent a single listing (p.14, lines 11-15). The processor then determines whether the identified records are actually equivalent (p. 14, lines 24-25). This is achieved by comparing each of the two normalized data records to the other normalized data record on a field-by-field basis (p. 15, lines 5-16). For example, the processor may compare the two records on a field-by-field basis and calculate a confidence value reflecting how many fields are identical (p. 15, lines 5-16). If the confidence level value is sufficiently high, the data records are deemed to be equivalent (p.15, lines 12-16).

When two records are deemed equivalent, the processor may create a final record in the uniform format (p. 16, line 16 - p. 17, line 6). For each defined field within the final record, data items from corresponding fields of the equivalent normalized records are grouped together and compared, and data items from one of the records is selected (p. 16, line14 - p. 17, line 6). Comparisons between data items in corresponding fields may be conducted based on, e.g.,

reliability rankings (p. 16, lines 14-26). The selected data item is copied to the corresponding field in the final record.

Kane discloses a system and method for utilizing data stored in one or more source databases to update data stored in a target database. Data records in the source databases are converted to a common format (col. 7, lines 13-20). The reformatted source records are then sequentially processed to update the target database (col. 9, lines 30-32).

Seilhamer discloses a relational database system for storing biomolecular sequence information in a manner that allows sequences to be catalogued and searched according to one or more protein function hierarchies (col. 2, lines 15-20). The hierarchies are provided to allow carefully tailored searches for sequences based upon a protein's biological or molecular function (col. 2, lines 15-22).

However, nowhere do Kane and Seilhamer, individually or in combination, teach or suggest, among others, "associating a time of the normalization process", which is "a function of time," with data assemblages undergoing normalization "to facilitate renormalization" of the assemblages "based on the time," as amended claims 1, 4, 7, 11, 15 and 18 now recite. As such, these claims, together with their dependent claims, are patentable over Kane in view of Seilhamer.

In view of the foregoing, each of claims 1-21, as amended, is believed to be in condition for allowance. Accordingly, reconsideration of these claims is requested and allowance of the application is earnestly solicited.

Respectfully,

By 

Alex L. Yip
Attorney for Applicant
Reg. No. 34,759
212-836-7363

Date: May 25, 2005